

GENE SIGNATURES FOR LIVER DISEASE AND USES THEREOF

Technical field

5 The present invention relates, in general terms, to biomarkers for liver disease. Disclosed herein are gene signatures for non-alcoholic fatty liver disease (NAFLD) and their uses for diagnosis and disease monitoring.

Background

10

Non-alcoholic fatty liver disease (NAFLD) is the most common chronic liver disease worldwide, affecting a quarter of the world population. NAFLD presents clinically as a spectrum of disease; a fraction of patients who initially exhibit simple steatosis will develop the progressive form of the disease, known as non-alcoholic steatohepatitis (NASH), which
15 may lead to liver fibrosis, cirrhosis and liver cancer. Current worldwide prevalence is about 25% for NAFLD and about 3% to 5% for NASH, with an increasing trend for both. With no pharmacological therapy currently approved, the socio-economic impact of this disease as well as the burden on the health system cannot be overestimated. Diagnosis of the disease still depends on liver biopsy as the gold standard, but biopsies are invasive and cannot be
20 performed routinely. Although imaging technologies like MRI and ultrasound-based fibroscan can also be used for diagnosis, they can only be performed at specialist and/or dedicated centres and are difficult to scale to allow population-wide screening. A non-invasive blood-based NAFLD diagnostic will make it easier to detect the disease early and begin intervention when the disease is still mild and more amenable to treatment. In addition,
25 a non-invasive diagnostic tool will make it easier to monitor patient response to treatment and thus facilitate the development of intervention strategies.

It would be desirable to overcome or ameliorate at least one of the above-described problems, or at least to provide a useful alternative.

30

Summary

Disclosed herein is a method of predicting the likelihood of a liver disease in a subject, the method comprising detecting in an extracellular vesicle sample from the subject the

expression of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, wherein an increase in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject is likely to have a liver disease.

Disclosed herein is a method of detecting a liver disease in a subject, the method comprising detecting in an extracellular vesicle sample from the subject the expression of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, wherein an increase in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject has a liver disease.

Disclosed herein is a method of treating a liver disease in a subject, the method comprising a) detecting in a sample from the subject the expression level of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, wherein an increase in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject has a liver disease, and b) treating the subject who has a liver disease.

Disclosed herein is a method of classifying a subject with NAFLD, comprising detecting in a sample from the subject the expression level of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, wherein an increase in the level of expression of the one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L* and *LIMS1*, or a value derived therefrom, as compared to a reference, indicates that the subject has non-alcoholic steatosis, and wherein an increase in the level of expression of the one or more genes selected from the group consisting of *CTXN1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, or a value derived therefrom, as compared to a reference, indicates that the subject has non-alcoholic steatohepatitis (NASH).

Disclosed herein is a method of monitoring a liver disease in a subject, the method comprising detecting in a sample from the subject the expression level of one or more genes

selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*.

Disclosed herein is a kit for detecting a liver disease in a subject, comprising reagents for
5 detecting in a sample from the subject the expression of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*.

Brief description of the drawings

10

Embodiments of the present invention will now be described, by way of non-limiting example, with reference to the drawings in which:

Figure 1 shows an outline of the time points for tissue collection.

15

Figure 2 shows mouse models of NAFLD and NASH.

Figure 3 shows fatty changes and fibrosis development in the mouse models.

20 **Figure 4** shows the generation of a high-quality RNA library from 500 ul of serum samples.

Figure 5 shows that human exosome RNAseq pathway analysis reveals progressive metabolic syndrome-related changes.

25 **Figure 6** shows steatosis exosome biomarker validation results for *CTXN1*, *CSNK1E* and *KHDC3L*.

Figure 7 shows validation results of steatosis polyA-enriched biomarkers *CSNK1E*, *KHDC3L* and *CTXN1*.

30

Figure 8 shows validation results of steatosis polyA-enriched biomarkers *NRGN* and *LIMS1*.

Figure 9 shows validation results of steatosis polyA-enriched biomarkers *CPE* and *EEF1A1*.

Figure 10 shows a combined biomarker heat map for steatosis with low fibrosis.

Figure 11 shows NASH exosome biomarker validation results for CTXN1, CLPS and
5 VIPR2.

Figure 12 shows validation results of NASH polyA-enriched biomarkers CTXN1, CLPS, VIPR2 and CIDEB.

10 **Figure 13** shows validation results of NASH polyA-enriched biomarkers MADCAM1 and HSD17B14.

Figure 14 shows validation results of NASH polyA-enriched biomarkers GAS1 and SLPI.

15 **Figure 15** shows validation results of NASH polyA-enriched biomarkers NAGS and CHCHD1.

Figure 16 shows a combined biomarker heat map for NASH with high fibrosis.

20 **Figure 17** shows the final gene signatures for steatosis and NASH.

Figure 18 shows an example of a workflow for biomarker detection or discovery using exosomal RNA derived from blood or serum.

25 **Detailed description**

The inventors have established gene signatures associated with two histopathological stages of non-alcoholic fatty liver disease (NAFLD), also known as metabolic dysfunction-associated steatotic liver disease (MASLD): the earlier and more benign stage of non-alcoholic steatosis (also known as metabolic dysfunction-associated steatotic liver, MASL),
30 and the later progressive stage of non-alcoholic steatohepatitis (NASH) (also known as metabolic dysfunction-associated steatohepatitis, MASH). The signatures are derived from exosomes isolated from blood serum. To identify high-confidence diagnostic markers, the inventors isolated and analysed RNA from exosomes from different NAFLD mouse models

and human patients. Disease stage-specific biomarkers that are common to both species were selected for further analysis. Biomarker validation was done on an independent set of human patient-derived samples by quantitative PCR (qPCR). A set of 16 genetic biomarkers which allow a precise disease stage diagnosis was established. Detection of the gene signatures can
5 be carried out using RNA extracted from liquid biopsy samples (e.g., from blood serum). The workflow of sample processing and analysis, which includes exosome isolation, RNA extraction, polyA enrichment and qPCR, can be performed in most diagnostic labs within a day. This has the potential to lower the cost of NAFLD testing and to open up NAFLD diagnostics to population-wide early screening, patient stratification and monitoring of
10 disease progression and treatment response.

The methods herein for identifying disease signatures using exosome-based RNA detection can be applied to other diseases, including liver diseases other than NAFLD. The methods can also be used to discover disease pathways. For instance, the inventors have found
15 through pathway analysis of the genes of this disclosure that metabolic syndrome-related changes associated with diabetes / insulin resistance and leaky gut syndrome may also be implicated in NAFLD. The inventors have also identified cardiovascular and bacterial changes associated with NAFLD.

Accordingly, this disclosure provides gene signatures for diagnosing, classifying and monitoring liver disease and in particular non-alcoholic fatty liver disease (NAFLD). The methods generally involve detecting gene expression of a panel of biomarkers from an extracellular vesicle sample from a subject, and diagnosing liver disease based on changes in gene expression. Methods herein may be used to predict the likelihood of a liver disease,
25 detect a liver disease, classify a liver disease as a particular subtype, monitor progression of a liver disease, and/or guide therapeutic intervention for a liver disease.

Disclosed herein is a method of predicting the likelihood of a liver disease in a subject, the method comprising detecting in an extracellular vesicle sample from the subject the
30 expression of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, wherein a change in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject is likely to have a liver disease.

Also disclosed herein is a method of detecting a liver disease in a subject, the method comprising detecting in an extracellular vesicle sample from the subject the expression of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*,
5 *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, wherein a change in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject has a liver disease.

10 “Liver disease”, as used herein, refers to a disease that affects liver function, and particularly to chronic metabolic, inflammatory and/or fibrotic diseases of the liver. The disease may be associated with metabolic syndrome in a subject. Examples of liver disease include non-alcoholic fatty liver disease (also known as metabolic dysfunction-associated steatotic liver disease), non-alcoholic steatosis (also known as metabolic dysfunction-associated steatotic
15 liver), non-alcoholic steatohepatitis (also known as metabolic dysfunction-associated steatohepatitis), liver fibrosis, and liver cirrhosis.

As used herein, an “extracellular vesicle (EV)” refers to a lipid bilayer-delimited nano- or micro-sized non-replicative particle secreted or shed from an eukaryotic cell. Extracellular
20 vesicles may range in size from about 10 nm to about 10 μ m. They carry a cargo of proteins, nucleic acids, lipids, metabolites and/or organelles from the parent cell, either within the vesicle or attached to the lipid membrane. EVs may be classified based on size, biochemical composition, cell of origin, biogenesis or condition under which they are released. Examples of EVs include but are not limited to ectosomes (released from the plasma membrane),
25 exosomes (derived from endosomes), apoptotic bodies (release by cells undergoing apoptosis), oncosomes (released by cancer cells) and migrasomes (released by migrating cells).

As used herein an “exosome” refers to an EV with a diameter of about 30 nm to about 200
30 nm. It is believed that materials endocytosed by a cell or other cellular components may be sorted into endosomal compartments, forming intraluminal vesicles within multivesicular bodies (MVBs). Fusion of MVBs with the plasma membrane may then release these vesicles into the extracellular environment as exosomes. Exosomes may be suitably identified by detecting known protein markers such as CD9, CD63, CD81, HSP70, HSP90, Alix, Tsg101,

Annexin (such as Annexin I, Annexin II or Annexin V), Flotillin (such as Flotillin-1 or Flotillin-2), and others.

5 A “subject” herein includes any organism with a liver, including but not limited to fish, birds such as chickens, ducks, guinea fowl, turkeys and quails, and mammals such as dogs, cats, pigs, cows, oxen, goats, horses, camels, shecp, rodents (such as mice, rats, guinea pigs and rabbits) and primates (including humans and non-human primates). In preferred embodiments, a subject herein is a human subject.

10 The term “sample” herein is used in its broadest sense. In one sense, it is meant to include a specimen or culture obtained from any source, including both biological and environmental sources. A “biological sample” includes within its scope a collection of similar fluids, cells, or tissues isolated from a biological source, such as a whole organism or *in vitro* culture. Samples include but are not limited to tissue biopsies, tissue resections, tissue aspirates,
15 swabs (e.g., buccal swabs), whole blood, plasma, serum, urine, saliva, cerebrospinal fluid, and cell cultures, and may be obtained using any suitable method known in the art. Archival tissues, such as those having treatment or outcome history may also be used for sample extraction. The sample may be pooled from multiple aliquots. Samples include untreated, treated, diluted and concentrated samples.

20 A “biological fluid” herein includes, but is not limited to, intravascular fluid (e.g., blood, plasma, serum, lymph), urine, saliva, sputum, cerebrospinal fluid, pleural fluid, fluid of the respiratory, intestinal, and genitourinary tracts, synovial fluid, vaginal secretion, tear fluid, pus, breast milk, semen, fluid from ascites, cyst or tumour, amniotic fluid, or combinations
25 thereof.

The terms “treating”, “treatment” and the like include relieving, reducing, alleviating, ameliorating or otherwise inhibiting the effects of the liver disease for at least a period of time. It is also to be understood that terms “treating”, “treatment” and the like do not imply
30 that the liver disease, or a symptom thereof, is permanently relieved, reduced, alleviated, ameliorated or otherwise inhibited and therefore also encompasses the temporary relief, reduction, alleviation, amelioration or otherwise inhibition of the disease, or of a symptom thereof.

As used herein a “therapeutically effective dose” or “effective dose” is an amount sufficient to effect desired clinical results (i.e., achieve therapeutic efficacy). A therapeutically effective dose can be administered in one or more administrations. For purposes of this disclosure, a therapeutically effective dose of a drug is an amount that is sufficient to palliate,
 5 ameliorate, stabilise, reverse, prevent, slow or delay the progression of a liver disease state, c.g., non-alcoholic steatosis or NASH.

Gene signatures

10 Gene signatures herein are based on one or a combination of genes selected from *CTXNI*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*. While changes in the expression of any one of the genes may be used to predict the likelihood of liver disease, the use of more genes (such as at least two genes, or a subset of the genes) is expected to increase the accuracy of
 15 disease prediction or diagnosis.

Sequences for the genes and their expression products may be accessed on known databases of genome, transcriptome, proteome or extracellular vesicle information, such as the GenBank or RefSeq databases maintained by the National Center for Biotechnology
 20 Information (NCBI), USA. For example, the following are NCBI GeneID identifiers for the human genes:

Human cortixin 1 (*CTXNI*): GeneID 404217

Human cascain kinase 1 epsilon (*CSNK1E*): GeneID 1454

Human eukaryotic translation elongation factor 1 alpha 1 (*EEF1A1*): GeneID 1915

25 Human neurogranin (*NRGN*): GeneID 4900

Human carboxypeptidase E (*CPE*): GeneID 1363

Human KH domain containing 3 like, subcortical maternal complex member (*KHDC3L*):
 GeneID 154288

Human LIM zinc finger domain containing 1 (*LIMS1*): GeneID 3987

30 Human mucosal vascular addressin cell adhesion molecule 1 (*MADCAM1*): GeneID 8174

Human coiled-coil-helix-coiled-coil-helix domain containing 1 (*CHCHD1*): GeneID
 118487

Human colipase (*CLPS*): GeneID 1208

Human hydroxysteroid 17-beta dehydrogenase 14 (*HSD17B14*): GeneID 51171

- Human secretory leukocyte peptidase inhibitor (*SLPI*): GeneID 6590
- Human cell death inducing DFFA like effector b (*CIDEB*): GeneID 27141
- Human N-acetylglutamate synthase (*NAGS*): GeneID 162417
- Human vasoactive intestinal peptide receptor 2 (*VIPR2*): GeneID 7434
- 5 Human growth arrest specific 1 (*GAS1*): GeneID 2619

In some embodiments, an increase in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject is likely to have a liver disease. Table 1 provides a list of possible combinations.

10

Table 1 – Possible Combinations for Predicting the Likelihood of Liver Disease

1	A method of predicting the likelihood of a liver disease in a subject, the method comprising detecting in an extracellular vesicle sample from the subject the expression of one or more genes selected from the group consisting of <i>CTXN1</i> , <i>CSNK1E</i> , <i>EEF1A</i> , <i>NRGN</i> , <i>CPE</i> , <i>KHDC3L</i> , <i>LIMS1</i> , <i>MADCAM1</i> , <i>CHCHD1</i> , <i>CLPS</i> , <i>HSD17B14</i> , <i>SLPI</i> , <i>CIDEB</i> , <i>NAGS</i> , <i>VIPR2</i> and <i>GAS1</i> , wherein an increase in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject is likely to have a liver disease.
2	The method of 1, wherein the panel of genes comprises 1 gene.
3	The method of 1, wherein the panel of genes comprises 2 genes.
4	The method of 1, wherein the panel of genes comprises 3 genes.
5	The method of 1, wherein the panel of genes comprises 4 genes.
6	The method of 1, wherein the panel of genes comprises 5 genes.
7	The method of 1, wherein the panel of genes comprises 6 genes.
8	The method of 1, wherein the panel of genes comprises 7 genes.
9	The method of 1, wherein the panel of genes comprises 8 genes.
10	The method of 1, wherein the panel of genes comprises 9 genes.
11	The method of 1, wherein the panel of genes comprises 10 genes.
12	The method of 1, wherein the panel of genes comprises 11 genes.
13	The method of 1, wherein the panel of genes comprises 12 genes.
14	The method of 1, wherein the panel of genes comprises 13 genes.
15	The method of 1, wherein the panel of genes comprises 14 genes.
16	The method of 1, wherein the panel of genes comprises 15 genes.
17	The method of 1, wherein the panel of genes comprises 16 genes.
18	The method of any one of 1-17, wherein the panel of genes comprises <i>CTXN1</i> .

19	The method of any one of 1-18, wherein the panel of genes comprises <i>CSNK1E</i> .
20	The method of any one of 1-19, wherein the panel of genes comprises <i>EEF1A</i> .
21	The method of any one of 1-20, wherein the panel of genes comprises <i>NRGN</i> .
22	The method of any one of 1-21, wherein the panel of genes comprises <i>CPE</i> .
23	The method of any one of 1-22, wherein the panel of genes comprises <i>KHDC3L</i> .
24	The method of any one of 1-23, wherein the panel of genes comprises <i>LIMS1</i> .
25	The method of any one of 1-24, wherein the panel of genes comprises <i>MADCAM1</i> .
26	The method of any one of 1-25, wherein the panel of genes comprises <i>CHCHD1</i> .
27	The method of any one of 1-26, wherein the panel of genes comprises <i>CLPS</i> .
28	The method of any one of 1-27, wherein the panel of genes comprises <i>HSD17B14</i> .
29	The method of any one of 1-28, wherein the panel of genes comprises <i>SLPI</i> .
30	The method of any one of 1-29, wherein the panel of genes comprises <i>CIDEB</i> .
31	The method of any one of 1-30, wherein the panel of genes comprises <i>NAGS</i> .
32	The method of any one of 1-31, wherein the panel of genes comprises <i>VIPR2</i> .
33	The method of any one of 1-32, wherein the panel of genes comprises <i>GAS1</i> .

In some embodiments, the liver disease is non-alcoholic fatty liver disease (NAFLD), non-alcoholic steatosis, or non-alcoholic steatohepatitis (NASH). The gene signatures can be used to generate a clinical assessment for diagnosing a subject with NAFLD, non-alcoholic steatosis or NASH; for classifying a NAFLD subject as one with steatosis or steatohepatitis; for monitoring NAFLD progression in a subject; or for determining if a subject is responding to an administered therapy.

In one embodiment, there is provided a method of predicting the likelihood of non-alcoholic steatosis in a subject, the method comprising detecting in an extracellular vesicle sample from the subject the expression of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L* and *LIMS1*, wherein a change in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject has or is likely to have non-alcoholic steatosis.

In some embodiments, an increase in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject is likely to have non-alcoholic steatosis. Table 2 provides a list of possible combinations.

Table 2 – Possible Combinations for Predicting the Likelihood of Non-alcoholic Steatosis

1	A method of predicting the likelihood of non-alcoholic steatosis in a subject, the method comprising detecting in an extracellular vesicle sample from the subject the expression of one or more genes selected from the group consisting of <i>CTXN1</i> , <i>CSNK1E</i> , <i>EEF1A</i> , <i>NRGN</i> , <i>CPE</i> , <i>KHDC3L</i> and <i>LIMS1</i> , wherein an increase in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject is likely to have non-alcoholic steatosis.
2	The method of 1, wherein the panel of genes comprises 1 gene.
3	The method of 1, wherein the panel of genes comprises 2 genes.
4	The method of 1, wherein the panel of genes comprises 3 genes.
5	The method of 1, wherein the panel of genes comprises 4 genes.
6	The method of 1, wherein the panel of genes comprises 5 genes.
7	The method of 1, wherein the panel of genes comprises 6 genes.
8	The method of 1, wherein the panel of genes comprises 7 genes.
9	The method of any one of 1-8, wherein the panel of genes comprises <i>CTXN1</i> .
10	The method of any one of 1-9, wherein the panel of genes comprises <i>CSNK1E</i> .
11	The method of any one of 1-10, wherein the panel of genes comprises <i>EEF1A</i> .
12	The method of any one of 1-11, wherein the panel of genes comprises <i>NRGN</i> .
13	The method of any one of 1-12, wherein the panel of genes comprises <i>CPE</i> .
14	The method of any one of 1-13, wherein the panel of genes comprises <i>KHDC3L</i> .
15	The method of any one of 1-14, wherein the panel of genes comprises <i>LIMS1</i> .

5 In one embodiment, there is provided a method of predicting the likelihood of NASH in a subject, the method comprising detecting in an extracellular vesicle sample from the subject the expression of one or more genes selected from the group consisting *CTXN1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, wherein a change in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject has or is likely to have NASH.

10

In some embodiments, an increase in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject is likely to have a liver disease. Table 3 provides a list of possible combinations.

Table 3 – Possible Combinations for Predicting the Likelihood of NASH

1	A method of predicting the likelihood of NASH in a subject, the method comprising detecting in an extracellular vesicle sample from the subject the expression of one or more genes selected from the group consisting <i>CTXN1</i> , <i>MADCAM1</i> , <i>CHCHD1</i> , <i>CLPS</i> , <i>HSD17B14</i> , <i>SLPI</i> , <i>CIDEB</i> , <i>NAGS</i> , <i>VIPR2</i> and <i>GAS1</i> , wherein an increase in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject is likely to have NASH.
2	The method of 1, wherein the panel of genes comprises 1 gene.
3	The method of 1, wherein the panel of genes comprises 2 genes.
4	The method of 1, wherein the panel of genes comprises 3 genes.
5	The method of 1, wherein the panel of genes comprises 4 genes.
6	The method of 1, wherein the panel of genes comprises 5 genes.
7	The method of 1, wherein the panel of genes comprises 6 genes.
8	The method of 1, wherein the panel of genes comprises 7 genes.
9	The method of 1, wherein the panel of genes comprises 8 genes.
10	The method of 1, wherein the panel of genes comprises 9 genes.
11	The method of 1, wherein the panel of genes comprises 10 genes.
12	The method of any one of 1-11, wherein the panel of genes comprises <i>CTXN1</i> .
13	The method of any one of 1-12, wherein the panel of genes comprises <i>MADCAM1</i> .
14	The method of any one of 1-13, wherein the panel of genes comprises <i>CHCHD1</i> .
15	The method of any one of 1-14, wherein the panel of genes comprises <i>CLPS</i> .
16	The method of any one of 1-15, wherein the panel of genes comprises <i>HSD17B14</i> .
17	The method of any one of 1-16, wherein the panel of genes comprises <i>SLPI</i> .
18	The method of any one of 1-17, wherein the panel of genes comprises <i>CIDEB</i> .
19	The method of any one of 1-18, wherein the panel of genes comprises <i>NAGS</i> .
20	The method of any one of 1-19, wherein the panel of genes comprises <i>VIPR2</i> .
21	The method of any one of 1-20, wherein the panel of genes comprises <i>GAS1</i> .

The expression of the *CTXN1* gene may be elevated in both non-alcoholic steatosis and NASH. Expression of *CTXN1* is generally higher in subjects with NASH than in subjects with non-alcoholic steatosis. The expression level of the *CTXN1* gene may be used in combination with one or more other biomarkers to determine if a subject has non-alcoholic steatosis or NASH. For instance, elevated expression of *CTXN1* in combination with one or more of *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L* and *LIMS1* indicates that a subject has non-alcoholic steatosis, and elevated expression of *CTXN1* in combination with one or more

of *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1* indicates that a subject has NASH.

An unchanged or decreased expression of one or more of the genes *CSNK1E*, *EEF1A*,
5 *NRGN*, *CPE*, *KHDC3L* and *LIMS1*, alongside an increase in the expression of one or more
of the genes *CTXN1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*,
VIPR2 and *GAS1*, relative to a reference, may indicate NAFLD progression from non-
alcoholic steatosis to NASH. Similarly, an increase in the expression of one or more of the
genes *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L* and *LIMS1* alongside an unchanged or
10 decreased expression of one or more of the genes *CTXN1*, *MADCAM1*, *CHCHD1*, *CLPS*,
HSD17B14, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, relative to a reference, may indicate
NAFLD regression from NASH to non-alcoholic steatosis.

The terms “increased” and “increase” are used herein to mean an increase by a statistically
15 significant amount. In some embodiments, the terms “increased” and “increase” can mean
an increase of at least 10% as compared to a reference level, for example an increase of at
least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about
60%, at least about 70%, at least about 80%, at least about 90% or up to and including a
100% increase or any increase between 10–100% as compared to a reference level, or at
20 least about a 2-fold, at least about a 3-fold, at least about a 4-fold, at least about a 5-fold or
at least about a 10-fold increase, or any increase between 2-fold and 10-fold or greater as
compared to a reference level.

The terms “decreased” and “decrease” are used herein to mean a decrease by a statistically
25 significant amount. In some embodiments, the terms “decreased” and “decrease” can mean
a decrease of at least 10% as compared to a reference level, for example a decrease of at
least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about
60%, at least about 70%, at least about 80%, at least about 90% or up to and including a
100% decrease or any decrease between 10–100% as compared to a reference level.

30

In some embodiments, an increase in gene expression is an increase of at least two-fold
relative to the reference.

The gene expression level in an EV sample is compared to a reference to determine overexpression or underexpression (or upregulation or downregulation). The reference may be the gene expression level in an EV sample from a subject of the same species without the liver disease, or an average gene expression level in EV samples from a population of
5 subjects of the same species (e.g., of varying ages, ethnic backgrounds and genders) without the liver disease. For monitoring a liver disease, the reference may also be the gene expression level in an EV sample from the same subject before the suspected onset of the liver disease, before the start of a treatment regimen, or at a different time-point during the course of the liver disease or during the course of treatment for the liver disease. The
10 reference values can be stored in a database and used as a reference in subsequent analyses.

The measured expression level of a gene may first be normalised before comparison with a reference. Normalisation can be used to control for unwanted biological variation. In a non-limiting example, biological variation can result from some feature of the patient or the
15 sample collection that is not relevant to the methods of the present disclosure, such as blood-exosome concentration due to high or low blood pressure, variations created by collecting samples at different times of the day and variations due to patient age or patient sex.

Normalisation can be performed using methods known in the art. In a non-limiting example,
20 normalisation can be achieved by dividing the measured expression level of a gene disclosed herein by a reference gene. Useful reference genes are genes that show a low variation in their expression level across a variety of different samples and patients. For example, a useful reference gene will show the same expression level in samples derived from subjects who have liver disease and in samples derived from subjects who do not have liver disease.
25 In another example, a useful reference gene will show the same expression level in samples derived from a subject with liver disease before treatment and in samples derived from a subject with liver disease after treatment. The variation in expression level can be quantified by different methods known in the art. For example, the variation in expression level of a gene can be quantified by calculating the coefficient of variation in the expression level of a
30 particular gene across a set of different samples.

Extracellular vesicle samples

Methods herein comprise detecting gene signatures in extracellular vesicle (EV) samples. In some embodiments the extracellular vesicle sample is an exosome sample. Exosomes are known to contain proteins as well as nucleic acids, including various DNA and RNA types such as mRNA (messenger RNA), hnRNA (heterogeneous nuclear RNA), miRNA (micro RNA), tRNA (transfer RNA), piRNA (piwi-interacting RNA), snRNA (small nuclear RNA), snoRNA (small nucleolar RNA), rRNA (ribosomal RNA), siRNA (short interfering RNA), shRNA (short hairpin RNA), and various classes of long non-coding RNA (lncRNA), including long intergenic non-coding RNA (lincRNA). Nucleic acids or proteins derived from exosomes and other extracellular vesicles can have a role as biomarkers for medical diagnosis, prognosis and therapy evaluation.

An EV sample herein can be isolated from any biological sample taken from a subject, e.g., a tissue sample from the liver, pancreas, stomach, gastrointestinal tract, lung, bladder, kidney, ovary, testis, skin, breast, prostate, brain, oesophagus, placenta, spleen, bone marrow, heart, pancreas, lymph node, and combinations thereof. The sample may be isolated from a liver biopsy. The sample may be a freshly harvested or an archived or stored sample.

In some embodiments, the EV sample is isolated from a biological fluid from a subject, e.g., from a liquid biopsy. Non-limiting examples of biological fluids include intravascular fluid (e.g., blood, plasma, serum, lymph), urine, saliva, sputum, cerebrospinal fluid, pleural fluid, nipple aspirate, fluid of the respiratory, intestinal, and genitourinary tracts, synovial fluid, vaginal secretion, tear fluid, pus, breast milk, semen, fluid from ascites, cyst or tumour, amniotic fluid, or combinations thereof. In one embodiment, the EV sample is isolated from blood, plasma or serum.

The volume of biological fluid used for EV isolation is at least 200 μ l, preferably 500 μ l or more. Larger volumes allow more EVs to be isolated and thus provide more EV content (e.g., nucleic acids and proteins) for detection of the gene signatures disclosed herein. In one embodiment, the EV sample is isolated from at least 500 μ l of the biological fluid. This allows sufficient mRNA to be extracted from the sample for polyA enrichment, which reduces signal noise when analysing gene expression based on RNA levels.

In some embodiments, the methods herein further comprise isolating the extracellular vesicle sample from the subject prior to detecting gene expression. Extracellular vesicles may be

concentrated or isolated from a biological sample using methods known in the art, e.g., using size exclusion chromatography, gel permeation chromatography, ion-exchange chromatography, density gradient centrifugation, differential centrifugation, nanomembrane ultrafiltration, immunoabsorbent capture, affinity purification, microfluidic separation, 5 commercially available extracellular vesicle purification kits, or combinations thereof.

The methods herein may involve a step of size selection or purification of exosomes from a sample from a subject or from an extracellular vesicle fraction, e.g., to provide an exosome sample comprising only exosomes of 30–150 nm diameter size.

10

The presence of EVs can be confirmed by detecting known markers that are specific to a class of EVs. For instance, known exosomal markers include but are not limited to protein markers of the endosomal pathway (e.g., CD9, CD63, CD81), heat shock proteins (e.g., HSP70) and MVB synthesis proteins (e.g., Alix and TSG101). These protein markers may 15 be detected via western blotting or other known means of detecting proteins. Transmission electron microscopy (TEM), dynamic light scattering (DLS), and NanoSight LM10 analysis can also be used to analyse the presence and purity of isolated EVs. For the methods herein, EV samples may be used without further purification to isolate cell type-specific EVs. However, EVs deriving from particular cells of origin may also be used, and such EVs may 20 be purified using cell type-specific markers on the EVs. For example, liver-specific EVs may be isolated using immunoprecipitation with antibodies against asialoglycoprotein receptor 1 (ASGR1).

Extracellular vesicles may be lysed to extract polypeptides and RNA. The lysis of EVs and 25 extraction of nucleic acids and/or polypeptides may be achieved with various methods known in the art. For example, RNA extraction may be achieved using RNA precipitation according to standard procedures and techniques known in the art. Such methods may utilise a nucleic acid-binding column to capture nucleic acids contained within the extracellular vesicles. Once bound, the nucleic acids can then be eluted using a buffer or solution suitable 30 to disrupt the interaction between the nucleic acids and the binding column, thereby eluting the nucleic acids. Alternatively, commercially available kits for purifying proteins and RNA from EVs may be used, e.g., Total Exosome RNA & Protein Isolation Kit from Invitrogen, exoRNeasy kit from Qiagen, SeraMir Exosome RNA Purification Kit from Systems Bioscience, or Exosomal RNA Isolation Kit from Norgen Biotek.

Gene detection

In some embodiments, gene expression levels in EV samples are detected by detecting the levels of a polypeptide and/or RNA expression product of the gene. Non-limiting examples of RNA expression products include messenger RNA (mRNA), heterogeneous nuclear RNA (hnRNA), microRNA (miRNA), small interfering RNA (siRNA), short hairpin RNA (shRNA), long non-coding RNA (lncRNA), piwi-interacting RNA (piRNA) and circular RNA (circRNA). Methods of detecting expression products such as RNA and proteins are well known to a person skilled in the art. Exemplary nucleic acid detection methods include blotting techniques (e.g., Northern blots), probe hybridisation-based methods, nucleic acid amplification-based methods and nucleic acid sequencing. Exemplary protein detection methods include gel electrophoresis (e.g., 2D electrophoresis), immunoassays, protein activity assays and mass spectrometry.

In some embodiments, the methods herein comprise detecting an RNA product of one or more genes selected from the group consisting *CTXN1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*. In some embodiments, the RNA expression product is an mRNA.

In some embodiments, the methods herein comprise detecting the mRNA level of one or more genes selected from the group consisting *CTXN1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*.

In some embodiments, extracted RNA is analysed directly without an amplification step. Direct analysis may be performed with different methods including, but not limited to, the NanoString technology described in Geiss et al. (*Nature Biotechnol*, 2008, 26(3): 317-325) and Goytain et al. (*Methods Mol Biol*, 2020, 2079:125-139), which are incorporated herein by reference.

In preferred embodiments, extracted RNA is reverse-transcribed into complementary DNA (cDNA) before further amplification for analysis. Such reverse transcription may be performed alone or in combination with an amplification step. One example of a method combining reverse transcription and amplification steps is reverse transcription polymerase

chain reaction (RT-PCR), which may be further modified to be quantitative, e.g., quantitative RT-PCR (qRT-PCR). Quantitative amplification will allow quantitative determination of RNA levels. Methods for reverse transcription with and without amplification are generally known in the art.

5

Nucleic acid amplification methods include, without limitation, polymerase chain reaction (PCR) and its variants such as *in situ* PCR and quantitative PCR, self-sustained sequence replication and its variants, and transcriptional amplification system and its variants (Kwoh et al., 1989), followed by the detection of the amplified molecules using techniques well known to those of skill in the art. Especially useful are those detection schemes designed for the detection of nucleic acid molecules if such molecules are present in very low numbers.

Complementary DNA may also be amplified using isothermal amplification, i.e., amplification of DNA that occurs at substantially the same temperature. The temperature may vary over the course of the amplification procedure so long as the temperature remains substantially the same. A number of isothermal amplification methods are known in the art, including but not limited to transcription mediated amplification (TMA), nucleic acid sequence-based amplification (NASBA), signal mediated amplification of RNA technology (SMART), strand displacement amplification (SDA), nicking enzyme amplification reaction (NEAR), rolling circle amplification (RCA), loop-mediated isothermal amplification (LAMP), isothermal multiple displacement amplification (MDA), helicase-dependent amplification (HDA), single primer isothermal amplification (SPIA), and cross-primed amplification (CPA).

The DNA that is amplified can be detected by detecting double-stranded DNA with double-stranded DNA detection reagents or specifically detecting amplified DNA sequences with sequence-specific probes. Any method of detecting amplified DNA is suitable for use in the present invention.

Methods for assessing RNA levels that do not require conversion of the RNA to cDNA are also known in the art and are suitable for use in the methods herein. For example, the nCounter™ Analysis system from Nanostring Technologies uses a digital molecular barcoding technology for multiplex measurement of RNA levels. The RNA sample is mixed with pairs of capture and reporter probes, tailored to each RNA sequence of interest. After

hybridisation and washing, probe-bound target nucleic acids are immobilised to a surface to detect the fluorescent barcodes of the reporter probes. This allows for up to 1000-plex measurement with high sensitivity and without amplification bias.

5 Extracted RNA can also be sequenced to determine gene expression. Sequencing methods can include, but are not limited to RNA-seq. In some aspects, RNA-seq comprises reverse transcribing at least one RNA molecule to produce at least one double-stranded complementary DNA molecule (dscDNA). Methods known in the art for creating a dscDNA library may be used. RNA-seq can further comprise appending sequencing adaptors to the
10 at least one dscDNA molecule, followed by amplification, and finally sequencing. Methods of sequencing known in the art, including sequencing by synthesis, can be used. The various RNA-seq methods known in the art may be used. Base abundances obtained using RNA-seq methods can be measured as read counts and normalized using methods known in the art. Gene abundances can also be reported in Reads Per Million (RPM) or Transcripts Per
15 Million (TPM).

Next-generation sequencing (NGS) or high-throughput sequencing may also be performed for DNA or RNA detection. These sequencing techniques allow for the identification of nucleic acids present in low or high abundance in a sample, or which are otherwise not detected by
20 more conventional hybridisation methods or a quantitative PCR method. NGS typically incorporates the addition of nucleotides followed by washing steps.

In some embodiments, the methods herein comprise detecting a polypeptide product (e.g., a protein or a protein fragment) of one or more genes selected from the group consisting
25 *CTXN1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*. The polypeptide may be a post-translationally modified polypeptide. The polypeptide may be detected using an antibody (e.g., an antibody labelled with a radiolabel, chromophore, fluorophore or enzyme), an antibody derivative (e.g., an antibody conjugated with a substrate or with the protein or ligand of a protein-ligand pair like biotin-streptavidin),
30 or an antibody fragment (e.g., a single-chain antibody, an isolated antibody hypervariable domain, etc.) which binds specifically with a polypeptide product of a gene disclosed herein. This detection may be performed as part of *in vitro* protein assay techniques well-known in the art, e.g., enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence staining. Alternatively, the known method

of surface plasmon resonance (SPR) may be used to detect the interaction between a polypeptide product and an antibody or antibody fragment that recognises it, and used to quantify the amount of the polypeptide in a sample.

- 5 Non-immunological methods may also be used to detect a polypeptide product of a gene disclosed herein. For instance, labelled aptamers (e.g., a radio-labeled, chromophore-labeled, fluorophore-labeled, or enzyme-labeled aptamer) may be used for polypeptide binding and detection. An aptamer refers to a nucleic acid that has a specific binding affinity for a target molecule. A suitable aptamer can be identified using any known method, including but not limited to the SELEX process, and prepared or synthesised in accordance with any known method, including chemical synthesis methods and enzymatic synthesis methods. Other binding agents for detecting a polypeptide may include, e.g., small molecules, lectins, ligand-binding receptors, affibodies, nanobodies, ankyrins, alternative antibody scaffolds (e.g. diabodies), imprinted polymers, avimers, peptidomimetics, 10 peptoids, peptide nucleic acids, threose nucleic acids, synthetic receptors, and modifications and fragments of these.

- A polypeptide product of a gene disclosed herein may also be quantified, via its proteotypic peptides, by known mass spectrometry techniques, non-limiting examples of which include: 20 electrospray ionization mass spectrometry (ESI-MS), ESI-MS/MS, ESI-MS/(MS)_n, matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS), surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS), desorption/ionization on silicon (DIOS), secondary ion mass spectrometry (SIMS), quadrupole time-of-flight (Q-TOF), tandem time-of-flight (TOF/TOF) technology, 25 atmospheric pressure chemical ionization mass spectrometry (APCI-MS), APCI-MS/MS, APCI-(MS)_n, atmospheric pressure photoionization mass spectrometry (APPI-MS), APPI-MS/MS, and APPI-(MS)_n, quadrupole mass spectrometry, Fourier transform mass spectrometry (FTMS), quantitative mass spectrometry, and ion trap mass spectrometry. Any of these techniques may be combined with selected reaction monitoring (SRM) to produce 30 more targeted quantitative measurements for a polypeptide product.

Probes and kits

Provided herein are kits for use in the methods herein to detect gene expression. The kits may comprise nucleic acids, such as oligonucleotides probes or primers, for detecting an RNA expression product (e.g., an mRNA) of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*,
5 *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*. Alternatively or additionally, the kits may comprise antigen binding molecules or antigen binding fragments thereof for detecting a polypeptide expression product (e.g., a protein or fragment thereof) of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*,
10 *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*.

Disclosed herein is a kit for detecting a liver disease in a subject, comprising reagents for detecting in a sample from the subject the expression of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*,
15 *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*.

In some embodiments the reagents in the kit comprise nucleic acids for detecting an mRNA product of the one or more genes. The nucleic acids may be, for example, oligonucleotide probes or primers capable of hybridising to an mRNA product of the one or more genes. The
20 design of hybridisation probes and primers are well-known in the art.

In some embodiments, the nucleic acids are oligonucleotide probes for detecting mRNA. Oligonucleotide probes herein are single-stranded DNA or RNA molecules designed to be substantially complementary to specific target nucleic acids (e.g., RNA), such that
25 hybridisation of the target sequence and the probes occur. This complementarity need not be perfect, so long as the probe is able to hybridise with its target sequence under suitable hybridisation conditions. For example, an oligonucleotide probe can comprise a sequence having a complementarity to a corresponding target sequence of at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%,
30 at least about 97%, at least about 99%, or 100%. In general, hybridisation will be influenced by the length of the probe, the pH, the temperature, the concentration of mono- and divalent cations, the proportion of G and C nucleotides in the region of hybridisation, and the possible presence of denaturants. Such variables also influence the time required for hybridisation.

The preferred conditions will therefore depend upon the particular application. Such empirical conditions, however, can be routinely determined without undue experimentation.

The oligonucleotide probe may comprise a detectable label. Detectable labels include, for example, chromogens, fluorophores, near-infrared dyes, chemiluminescent molecules, bioluminescent molecules, colloidal metal particles (such as gold or silver particles), lanthanide ions (e.g., Eu^{3+}), semiconductor nanocrystals (e.g., quantum dots), radioisotopes, epitopes, enzymes, colored beads (such as glass or plastic beads). Detectable labels also include barcodes such as molecular barcodes and fluorescent barcodes. Alternatively or additionally, the probe may contain a sequence for further amplification and detection, e.g., by sequencing.

In some embodiments, the nucleic acids in the kit are primers for reverse transcription of RNA to cDNA and/or for amplifying the RNA or cDNA product (e.g., primers for PCR amplification or isothermal amplification).

In certain embodiments, more than one oligonucleotide probe or primer is used per target sequence. The probes or primers may target overlapping sections or separate sections of the target sequence. That is, two, three, four or more probes or primers may be used to build in a redundancy for a particular target.

Nucleic acid probes or primers in the kit may comprise an affinity tag for separating target nucleic acids hybridised to the probe or primer. As used herein, the term “affinity tag” refers to a component of a multi-component complex, wherein the components of the multi-component complex specifically interact with or bind to each other. For example, an affinity tag can include biotin that can bind streptavidin. Other examples of multiple-component affinity tag complexes include ligands and their receptors; binding proteins/peptides, including maltose/maltose binding protein (MBP), calcium/calcium binding protein/peptide (CBP); antigen-antibody, including epitope tags, such as c-MYC, HA, VSV-G, HSV, V5, and FLAG Tag™, and their corresponding anti-epitope antibodies; haptens, for example, dinitrophenyl and digoxigenin, and their corresponding antibodies; aptamers and their corresponding targets; fluorophores and anti-fluorophore antibodies; and the like.

Nucleic acids herein may be immobilised on a solid substrate to form an array. Suitable substrates include plastics (e.g., polytetrafluoroethylene, polystyrene, acrylic polymers), ceramics, silicon, glass, ferro- or paramagnetic materials, titanium dioxide, latex, cross-linked dextrans such as Sepharose, and cellulose. Signal detection from an array may be performed using a chip or plate reader. Alternatively, when the array is made using a mixture of individually addressable kinds of labelled particles, the reaction may be detected using flow cytometry.

Kits herein may comprise reagents for generating a detectable signal upon probe or primer hybridisation to a target nucleic acid. For example, for probes conjugated to an enzyme, the kit may contain reagents and substrates for the enzyme to generate a detectable signal. The kit may also comprise double-stranded DNA detection reagents, such as ethidium bromide, PicogreenTM (Life Technologies) and SYBRTM Green (Life Technologies), among others.

Where the nucleic acid probes are primers, the kit may comprise an enzyme for performing the reverse transcription (e.g., a reverse transcriptase) and/or DNA-dependent DNA polymerisation (e.g., a DNA polymerase). The reagents may also comprise deoxynucleotide triphosphates (dNTPs), such as deoxyadenosine triphosphate (dATP), deoxyguanosine triphosphate (dGTP), deoxythymidine triphosphate (dTTP), deoxyuridine triphosphate (dUTP), deoxycytidine triphosphate (dCTP), deoxyinosine triphosphate (dITP), and deoxyxanthosine triphosphate (dXTP). The enzyme and/or deoxynucleotide triphosphates are preferably present in a combination and an amount sufficient to generate a DNA copy of the RNA template and to amplify the DNA copy into a detectable amount.

Some or all of the reagents in the kit may be provided together as a mix in dried form (e.g., a powder) and are reconstituted with a liquid prior to use. The reagents may be dried through lyophilisation or other methods known in the art. The kit may further comprise an aqueous buffer for reconstituting the dried components. The buffer may further comprise a magnesium salt (e.g., MgCl₂) and an organic solvent, such as dimethyl sulfoxide.

In some embodiments, the kit further comprises devices and reagents for isolating an exosome fraction from the biological fluid sample. The kit may comprise reagents for precipitating extracellular vesicles from the biological fluid, e.g., chaotropic agents like guanidinium thiocyanate. The kit may also comprise polymers (for example and without

- limitation: polyethylene glycol, dextran, dextran sulphate, dextran acetate, polyvinyl alcohol, polyvinyl acetate and polyvinyl sulphate) suitable for use in chromatography to separate exosomes from the fluid sample, e.g., size exclusion chromatography, gel permeation chromatography or ion-exchange chromatography. The polymers may have
- 5 molecular weights between 1 to 35 kDa. The polymers may be in the form of particles (e.g., microbeads). The polymers may further comprise polypeptides (e.g., antibodies, antibody fragments, antibody-like scaffolds) or nucleic acids (e.g., aptamers) that bind to moieties specific to exosomes (e.g., exosome-specific proteins or glycans) to enable exosome
- 10 may be covalently bound to the polymers. Alternatively, the kit may comprise magnetic or paramagnetic particles with conjugated exosome-binding polypeptides or nucleic acids. The kit may further comprise columns compatible with ultracentrifugation, for packing the polymers or magnetic particles to facilitate chromatographic separation.
- 15 In some embodiments, the kit further comprises devices and reagents for isolating RNA (e.g., mRNA) from the exosome fraction. The reagents may include chemical agents for lysing the exosomes (e.g., chaotropes like guanidinium thiocyanate, or detergents like Tween or Triton X-100) and precipitating proteins and/or nucleic acids. The devices may include ultracentrifugation columns containing nanoporous membranes for filtering out and
- 20 washing the precipitated nucleic acids. The kit may also include a DNase (e.g., DNase I) for removing DNA in the nucleic acid isolate. The kit may further contain oligo deoxythymidine (oligo dT) nucleic acids conjugated to a membrane, filter or particle (e.g., a polymeric or magnetic particle) for binding to and retaining mRNA from the exosome sample or lysate.
- 25 Disclosed herein is a composition comprising a) an extracellular vesicle sample obtained from a subject with a liver disease, and b) one or more nucleic acid probes for detecting the expression of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*.

30

Methods of diagnosis and monitoring

Disclosed herein is a method of predicting the likelihood of a liver disease in a subject, the method comprising detecting in an extracellular vesicle sample from the subject the

expression of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, wherein an increase in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject is likely to have a liver disease.

Disclosed herein is a method of predicting the likelihood of liver fibrosis in a subject, the method comprising detecting in an extracellular vesicle sample from the subject the expression of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, wherein an increase in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject is likely to have liver fibrosis.

The likelihood may be a numerical value, for example, a likelihood score or percentage. This value is empirically derived and will change depending on the data, cohort of the subject population, type of liver disease, number of biomarkers used for evaluation, reference used for comparison, etc. The value indicating likelihood of having a liver disease may be a numerical score, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, and so on, or some fraction thereof. Alternatively, the value indicating likelihood of having a liver disease may be a percentage, e.g., 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or up to and including 100%, and any value or fractional value in between.

Also disclosed herein is a method of detecting a liver disease in a subject, the method comprising detecting in an extracellular vesicle sample from the subject the expression of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, wherein an increase in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject has a liver disease.

In some embodiments, the liver disease is non-alcoholic fatty liver disease (NAFLD), non-alcoholic steatosis, or non-alcoholic steatohepatitis (NASH).

Disclosed herein is a method of classifying a subject with NAFLD, comprising detecting in a sample from the subject the expression level of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*,
5 *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, wherein an increase in the level of expression of the one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L* and *LIMS1*, or a value derived therefrom, as compared to a reference, indicates that the subject has non-alcoholic steatosis, and wherein an increase in the level of expression of the one or more genes selected from the
10 group consisting of *CTXN1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, or a value derived therefrom, as compared to a reference, indicates that the subject has non-alcoholic steatohepatitis (NASH).

Disclosed herein is a method of staging a sample from a subject with NAFLD, comprising
15 detecting in the sample the expression level of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, wherein an increase in the level of expression of the one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L* and *LIMS1*, or a value derived therefrom,
20 as compared to a reference, indicates that the subject has non-alcoholic steatosis, and wherein an increase in the level of expression of the one or more genes selected from the group consisting of *CTXN1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, or a value derived therefrom, as compared to a reference, indicates that the subject has non-alcoholic steatohepatitis (NASH).

25 Disclosed herein is a method of detecting liver fibrosis in a subject with NAFLD, comprising detecting in a sample from the subject the expression level of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*,
30 wherein an increase in the level of expression of the one or more genes, or a value derived therefrom, indicates that the subject has liver fibrosis.

In one embodiment, the value derived from the level of the one or more genes is a probability. The probability may be derived from a logistic regression model, where the level

of the one or more genes is the predictor of the logistic regression model. The value may be compared to a reference which may, for example, be a cutoff designated at 0.5.

5 Methods herein may be further complemented by other clinical diagnostic methods to provide a more comprehensive clinical assessment. Additional diagnostic procedures may include liver imaging (e.g., transient elastography, magnetic resonance elastography) and histopathological analyses using liver biopsy samples.

10 Methods herein may be performed on multiple occasions over a period of time (i.e., as part of a longitudinal screen) to monitor the progress of a liver disease or to monitor the response of a subject to a therapeutic intervention. Methods herein may also be used to guide therapeutic intervention for liver disease in a subject. For example, drug dosages or treatment regimens may be suitably adjusted or a different course of treatment may be selected depending on the diagnosis following a course of treatment.

15 Accordingly, disclosed herein is a method of monitoring a liver disease in a subject, the method comprising detecting in a sample from the subject the expression level of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*.

20 In some embodiments, the subject is undergoing or has undergone treatment for the liver disease. In some embodiments the liver disease is NAFLD. The subject may be receiving or has received treatment for NAFLD, e.g., lifestyle modification (e.g., weight loss or dietary changes) and/or pharmacological interventions.

25 An unchanged or decreased expression of one or more of the genes *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L* and *LIMS1*, alongside an increase in the expression of one or more of the genes *CTXN1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, compared to a reference, may indicate NAFLD progression from non-alcoholic steatosis to NASH. Similarly, an increase in the expression of one or more of the genes *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L* and *LIMS1*, alongside an unchanged or decreased expression of one or more of the genes *CTXN1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, compared to a reference, may indicate

NAFLD regression from NASH to non-alcoholic steatosis.

Disclosed herein is a method of treating a liver disease in a subject, the method comprising
 a) detecting in a sample from the subject the expression level of one or more genes selected
 5 from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*,
MADCAM1, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*,
 wherein an increase in the level of expression of the one or more genes, or a value derived
 therefrom, as compared to a reference indicates that the subject has a liver disease, and b)
 treating the subject who has a liver disease.

10

In some embodiments, the treatment comprises administering to the subject of an effective
 dose of one or a combination of NAFLD therapeutic agents. NAFLD therapeutic agents
 herein are any drugs or pharmaceutical compositions effective to treat NAFLD, non-
 alcoholic steatosis or NASH. Non-limiting examples of NAFLD therapeutic agents include
 15 insulin sensitisers (e.g., thiazolidinediones, metformin), antioxidants (e.g., vitamin E), lipid-
 lowering drugs (e.g., statins, ezetimibe), pentoxifylline, farnesoid X receptor agonists (e.g.,
 cilofexor, tropifexor, obeticholic acid, Px-104, EDP-305), peroxisome proliferator-activated
 receptor (PPAR) agonists (e.g., pioglitazone, PXL065, lanifibranor, saroglitazar, seladelpar,
 clafibranor), angiotensin receptor blockers (e.g., telmisartan, losartan), n-3 polyunsaturated
 20 fatty acids (e.g., eicosapentaenoic acid (EPA), docosahexaenoic acid (DHA)), sodium-
 glucose cotransporter 2 (SGLT2) inhibitors (e.g., dapagliflozin, canagliflozin,
 empagliflozin), thyroid hormone receptor beta agonists (e.g., resmetirom), and glucagon-
 like peptide-1 receptor (GLP-1R) agonists (e.g., semaglutide, exenatide, liraglutide,
 dulaglutide, lixisenatide). In other embodiments, the treatment comprises lifestyle
 25 modification, such as the prescription of specific diet or exercise regimens.

Disclosed herein is an NAFLD therapeutic agent for use in treating a liver disease in a
 subject, wherein the expression of one or more genes selected from the group consisting of
CTXN1, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*,
 30 *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1* has been detected in an extracellular
 vesicle sample obtained from the subject, and wherein an increase in the level of expression
 of the one or more genes, or a value derived therefrom, as compared to a reference, indicated
 that the subject has or is likely to have a liver disease.

Disclosed herein is the use of an NAFLD therapeutic agent in the manufacture of a medicament for treating a liver disease in a subject, wherein the expression of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1* has been detected in an extracellular vesicle sample obtained from the subject, and wherein an increase in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference, indicated that the subject has or is likely to have a liver disease.

The biomarkers described herein may guide or assist a physician in deciding a treatment path, for example, whether to implement procedures such as surgery, treat with drug therapy, or employ a watchful waiting approach. As with the other methods described herein, the comparisons made in the methods of monitoring progression/regression of liver disease in a subject may be carried out using various techniques, including simple comparisons, one or more statistical analyses, mathematical models (algorithms) and combinations thereof. The results of the method may be used along with other methods (or the results thereof) useful in the clinical monitoring of progression/regression of liver disease in a subject.

As used herein, "and/or" refers to and encompasses any and all possible combinations of one or more of the associated listed items, as well as the lack of combinations when interpreted in the alternative (or).

As used in this application, the singular form "a," "an," and "the" include plural references unless the context clearly dictates otherwise. For example, the term "an agent" includes a plurality of agents, including mixtures thereof.

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

Throughout this specification and the claims which follow, unless the context requires otherwise, the phrase "consisting essentially of", and variations such as "consists essentially of" will be understood to indicate that the recited element(s) is/are essential i.e. necessary

elements of the invention. The phrase allows for the presence of other non-recited elements which do not materially affect the characteristics of the invention but excludes additional unspecified elements which would affect the basic and novel characteristics of the method defined.

5

The reference in this specification to any prior publication (or information derived from it), or to any matter which is known, is not, and should not be taken as an acknowledgment or admission or any form of suggestion that that prior publication (or information derived from it) or known matter forms part of the common general knowledge in the field of endeavour to which this specification relates.

10

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications, which fall within the spirit and scope. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations of any two or more of said steps or features.

15

Unless otherwise defined, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs.

20

Certain embodiments of the invention will now be described with reference to the following examples which are intended for the purpose of illustration only and are not intended to limit the scope of the generality hereinbefore described.

25

Examples

Example 1: Preparation of RNA sequencing library from exosome samples

30

Different animal models of NAFLD and liver damage were established. Due to the complex interactions of different tissue types in NAFLD, *in vivo* mouse models that recapitulate the human disease are currently the best option to study NAFLD. Different mouse models based on diet changes or genetic engineering have been developed to reflect different aspects of

NAFLD and/or NASH (Fig. 2). Mice fed a “Western diet” exhibit obesity, insulin resistance, dyslipidemia, hyperglycemia and NAFLD. Western diets are considered high in both fat and fructose. Prolonged feeding also leads to patterns of inflammation, fibrosis, endoplasmic reticulum stress and lipoapoptosis with increased steatohepatitis, including hepatocyte ballooning. However, this model is slow-progressing and takes half a year to reach higher fibrosis levels. In contrast to the Western diet model, the choline-deficient L-amino acid defined high fat diet (CDHFD) progresses very fast to advanced steatohepatitis, reaching advanced fibrosis in 8 weeks. Both models are also used to discover and to investigate therapeutic interventions.

10

For establishing disease stage-specific signatures in mice, serum was collected from mice exposed for 1, 6 or 8 weeks to CDHFD; 8, 14, 20 or 26 weeks to “Western Diet” (WD); or respective controls exposed to normal chow for the indicated times (Fig. 1). Exosomes were isolated from the serum by ExoQuick precipitation solution (SBI, polymer based reagent). From the precipitated exosomes, RNA was isolated and used for generating Illumina-based next-generation sequencing compatible RNA libraries. The same isolation and analysis techniques were used for NAFLD patient-derived serum.

15

200 μ l of serum from disease stage-diagnosed patients from a local NAFLD cohort were used for analysis. For exosome isolation and RNA library generation this is the minimum volume required for library generation. Western blot was used to verify precipitation of exosomes by ExoQuick, using the exosome markers HSP70, CD63, and CD9.

20

RNA from precipitated exosomes was isolated using a modified protocol based on Total Exosome RNA Isolation Kit (Invitrogen). Briefly, the exosomes were resuspended in 1x PBS in an RNase-free tube to a volume of 200 μ l. An equivalent volume of 2x Denaturing Solution was added to the PBS suspension and thoroughly mixed, then incubated on ice for 5 min. A volume of acid-phenol:chloroform equivalent to the total volume of PBS and Denaturing Solution was added to each sample and vortexed for 30–60 seconds. The mixture was then centrifuged at $\geq 10,000$ g and room temperature for 5 min at to separate the aqueous and organic phases. The upper (aqueous) phase was carefully transferred to a fresh tube, and 1.5x volume of isopropanol and 1 μ L of Pellet Paint® Co-Precipitant (69049 Millipore, Merck) was added to the aqueous fraction and left overnight at -20°C. The mixture was centrifuged at $\geq 12,000$ g and 4°C for 40 min. The pellet was washed with 200–500 μ l of

25

30

fresh-prepared 70% ethanol, then centrifuged for 20 min at $\geq 12,000\ g$ and 4°C , and left to air dry for 10 min. Finally, the pellet was resuspended in 10–20 μL of RNase-free water and stored at -80°C until use.

5 The amount of RNA obtained from exosomes was used to prepare high quality Next-
Generation Sequencing (NGS) libraries for sequencing (TrueSeq® RNA Sample
Preparation v2 Guide, Illumina). The quality of the RNA libraries for NGS sequencing was
checked by Agilent, showing a good quality for mouse as well as human derived serum
exosome RNA (Fig. 4). Pooled paired-end Illumina-based deep sequencing was then
10 performed.

For the mouse models, sequencing was performed on exosome RNA derived from mice
which were exposed to the CDHFD for 1, 6, 8 weeks, or exposed to the WD for 8, 14, 20,
26 weeks. First, samples underwent a pre-alignment QA/QC, which showed a quality score
15 of $> Q30$ for the sequences. A Q30 means that the probability of an incorrect base call is 1
in 1000 and this translates to a base call accuracy of 99.9%. These sequencing results support
the high quality of RNA we isolated (Fig. 4). Importantly, high quality RNA can be obtained
independent of disease stage and animal model. Unaligned reads were aligned using STAR
2.6. Mouse samples were aligned to mouse genome mm10 reference. Aligned reads were
20 then quantified to the annotation model (mm10_gencode_M22_v2.pannot.transcripts --
gene_level_annotation). Gene counts were normalized by CPM and 0.001 was added. Gene
set analysis was done on normalized counts, including gene set enrichment and pathway
enrichment analysis. For this whole framework Partek Flow software solution was used.

25 Human serum samples were similarly processed as described for mice. Samples underwent
a pre-alignment QA/QC, which showed a quality score of $> Q30$ for our sequences. For
human sample, unaligned reads were aligned using STAR 2.6 to human reference genome
hg38. A more conservative alignment method was chosen to increase the confidence in
potential biomarkers. Aligned reads were quantified to the annotation model
30 (hg38_refseq_v92_1911_01_v2.pannot.transcripts -- gene_level). Gene counts were
normalised by CPM and 0.001 was added. Gene set analysis was done on normalised counts,
including gene set enrichment and pathway enrichment analysis.

Example 2: Identification and validation of exosomal RNA signatures for steatosis and NASH

Pathway analysis of mouse- and human-derived samples showed that the mRNA content of isolated exosomes reflects important gene expression changes associated with the metabolic syndrome (Fig. 5). For the Western Diet model, there was significant dysregulation of pathways beginning at 6 weeks diet exposure and overlapping to the end of 26 weeks. Some pathways are known to play a role in diet as well as in the metabolic syndrome, e.g., mTOR signaling pathway, MAPK signaling pathway, Type II diabetes mellitus, insulin signaling pathway. At the later stage of disease when there is fibrosis, NASH-related pathways are also enriched, e.g., TGF-beta signaling pathway, NF-kappa B signaling pathway, PI3K-Akt signaling pathway, ABC transporters. The pathway analysis supports the notion that the WD model recapitulates essential changes of progressive NAFLD. The same approach was applied to the data from the CDHFD model, where NAFLD-related pathway activation was also observed.

Overlaps in the exosome RNA signature from human NASH, WD 26 weeks and CDHFD 8 weeks, which all represent the NASH disease stage with advanced fibrosis, were identified. There was an overlap in dysregulated pathways associated with the disease. Many pathways are linked to cancer development, with links also seen to metabolic pathways and lipotoxicity. Species overlap at the pathway level suggests that mice can be used for human biomarker discovery, but there are some differences in data between mouse and man. For example, looking at the MAPK signaling pathway, there is in both species dysregulation of the pathway, but many more genes are upregulated in humans compared to mice.

For the initial set of samples from 3 healthy control patients, 3 patients with steatosis, and 3 patients with NASH, the full pathological report on the patients were available for steatosis grade, lobular inflammation, ballooning, and fibrosis grade. First, by looking at the broader pathway level (Fig. 5), liver-related and also cardiovascular and bacterial changes were seen to be associated with the disease. This fits with the notion that NAFLD is the liver manifestation of a broader metabolic syndrome. At the simple steatosis stage, there was already insulin resistance and enrichment of cysteine and methionine metabolism pathways. Insulin resistance is an early sign of the metabolic syndrome and dysregulated hepatic methionine metabolism drives NAFLD. When looking at enriched pathways common to

simple steatosis and NASH samples, insulin signaling is again implicated, along with such classic NAFLD-related pathways like MAPK, TGF- β and Hippo signalling. Specific to the NASH stage was an enrichment for Wnt and mTOR signalling. There were also cardiovascular-related changes associated with metabolic syndrome. Furthermore, pathways related to bacterial invasion and *E. coli* infection were enriched. NAFLD in the later stages is associated with leaky gut syndrome caused by increased permeability of the intestine, which allows lipopolysaccharide components of bacteria as well as other bacterial structural compounds to enter the entero-hepatic circulation, thus triggering stress responses in the liver. These results clearly show the power of exosome RNA biomarkers for identifying NAFLD-related pathways.

Fourteen biomarkers derived from the initial human samples were validated on an expanded patient sample set (8 healthy, 6 steatotic and 6 NASH patients) using qPCR. 200 μ l serum input was used. Several biomarkers validated in this cohort are directly associated with lipid or metabolic pathways. Samples that could not be tested (due to insufficient RNA from small serum input) are indicated with a white box and "NA". The isolated exosomes can come from any cell and organ, but most of the biomarkers identified using these exosomes show a clear disease stage-specific linkage. One biomarker does not show expression in the liver but instead codes for a protein only expressed in pancreatic acinar cells, an organ affected by metabolic syndrome.

To increase the accuracy of the diagnostic panel, additional human serum samples were obtained from American and European cohorts. Ten steatosis markers were first tested on nine serum samples from healthy American subjects. For exosome RNA isolation, 500 μ l of serum was processed as previously described. Six of the previously-identified biomarkers passed qPCR analysis (Fig. 6). Next, ten NASH markers were tested and six passed qPCR analysis (Fig. 11).

Since the method for biomarker discovery is focused on detecting protein coding RNA in exosomes, polyA enrichment of RNA was next performed on samples derived from 500 μ l of serum. Nine steatosis serum samples from German subjects, subdivided into fibrosis score 0 and 1 (Fig. 14 and 15), were used. The steatosis biomarker panel was tested on the nine healthy and nine steatosis polyA samples by qPCR. Three of the biomarkers scored significantly in the steatosis group with fibrotic score 1, with two showing a clear tendency.

However, none of the biomarkers scored significantly in steatosis samples with no fibrosis (Fig. 7). The biomarker panel was also tested on nine healthy and nine NASH polyA samples from American subjects. 3 markers strongly scored in the NASH sample but not in the healthy controls (Fig. 12). The fully validated biomarkers show a good NAFLD disease stage-specific diagnostic potential.

In-depth data mining was next performed on the mouse WD 20-week data set for steatosis and CDHFD data set for NASH (Fig. 16). For WD 20-week there were 5 mice with a fibrotic score of 1 and 3 mice with a fibrotic score of 2 (which resembles the human advanced steatosis stage). As control 5 mice showing no fibrosis after 20 weeks of normal chow were used. The exosome RNA sequences were aligned to mouse genome mm10 reference. Aligned reads were then quantified to the annotation model (mm10_gencode_M22_v2.pannot.transcripts -- gene_level_annotation). Gene counts were normalized by CPM and 0.001 was added. The normalized gene_level count matrix was then used as input to AltAnalyze version 2.1.4. (altanalyze.org / <https://github.com/nsalomonis/altanalyze>). Especially the MarkerFinder function was used to identify signature marker for the steatosis stage. A number of potential specific markers was identified and further filtered for human homologues. Four additional biomarkers were validated in the human steatosis set (Figs. 8-10). The markers specifically scored for steatosis with a fibrosis score of 1. In line with this the normalised counts for the mouse RNAseq data set show a fibrosis grade dependent significant increase. Interestingly several of the biomarker gene products are linked to post-translational acetylation and it was described previously that the protein acetylation pattern in fatty livers is significantly different from healthy livers.

Similar to the WD 20-week data mining, data from 5 mice which were exposed for 8 weeks to the choline-deficient L-amino acid defined high fat diet and 7 respective control animals were used to identify additional NASH-specific markers. All control animals showed a fibrosis score of 0, 2 of the CDHFD mice showed a fibrosis score of 3, and 3 mice a score of 4. The exosome RNA sequences were aligned to mouse genome mm10 reference. Aligned reads were then quantified to the annotation model (mm10_gencode_M22_v2.pannot.transcripts -- gene_level_annotation). Gene counts were normalized by CPM and 0.001 was added. The normalized gene_level count matrix was then used as input to AltAnalyze version 2.1.4. (altanalyze.org /

https://github.com/nsalomonis/altanalyze). Especially the MarkerFinder function was used to identify signatures for the NASH stage. Potential specific markers were filtered for genes with human homologues. 7 additional biomarkers were validated in the human NASH set (Figs. 13-16). The gene products of the additional validated biomarkers are linked to the hedgehog signaling pathway, mitochondrial ribosomes, and urea cycle.

In summary, re-examination and in-depth data mining of the NAFLD mouse model-derived exosome RNA expression data added 11 validated NAFLD diagnostic biomarkers. These biomarkers can be especially robust as they are conserved in mice and humans.

A set of 16 markers are established which in combination allow a precise disease stage diagnosis of NAFLD. As input a minimum of 500 ul of serum is expected, which is easily obtainable from standard blood-draw samples. Exosomes are simply isolated by precipitation. The whole process from exosome isolation to RNA extraction and analysis can be implemented in most diagnostic labs. The processing time for a sample can be less than a day.

Table 4 – Sequences of primers used for RT-PCR.

Gene name	Sequence (5' → 3') (Forward)	Sequence (5' → 3') (Reverse)
<i>CTXNI</i>	GGTCGAAGCCAGGGGCG (SEQ ID NO: 1)	TCGACGGCGGCAGGGGC (SEQ ID NO: 2)
<i>CSNK1E</i>	AAACAAGAACCTGACCGGC (SEQ ID NO: 3)	ACGGCTTTGCTCAATGCCC (SEQ ID NO: 4)
<i>EEF1A1</i>	TGCTGAGATGGGAAAGGG (SEQ ID NO: 5)	ACAAGGAGATATCAATGGTG (SEQ ID NO: 6)
<i>NRGN</i>	TGCTGCACCGAGAACGCC (SEQ ID NO: 7)	AAAACTCGCCTGGATTTTGG (SEQ ID NO: 8)
<i>CPE</i>	ACAAAGCTTGCTCCTGAGAC (SEQ ID NO: 9)	TGAGCACTACCACTCCGCG (SEQ ID NO: 10)
<i>KHDC3L</i>	AGGAGGACACAATCAAGATG (SEQ ID NO: 11)	TTCCTGAGCCTTTCGCCTG (SEQ ID NO: 12)
<i>LIMS1</i>	AAAGCCATGAATAACAGCTGG (SEQ ID NO: 13)	ACACAGGTGTCTCCCAGC (SEQ ID NO: 14)

<i>MADCAMI</i>	TCCTCCTCGGCCAGTCC (SEQ ID NO: 15)	AAGGCCACGGCACCACC (SEQ ID NO: 16)
<i>CHCHDI</i>	TGTGCTGAAGCCCAATAAAC (SEQ ID NO: 17)	ATGCAAGTCGCCTCGCCC (SEQ ID NO: 18)
<i>CLPS</i>	TCAAGTGCCTGGGCCTG (SEQ ID NO: 19)	ATAGAGCGTCTTGACAGAGC (SEQ ID NO: 20)
<i>HSD17B14</i>	TCGTGCGCGCCTTCGTG (SEQ ID NO: 21)	TCCTGCTCCAGGGCCCG (SEQ ID NO: 22)
<i>SLPI</i>	AAACCCAACAAGGAGGAAGC (SEQ ID NO: 23)	ATCCATCTCACAGAAATTGGG (SEQ ID NO: 24)
<i>CIDEB</i>	TTCCGTGTCTGTGATCACAAG (SEQ ID NO: 25)	TCCAATGCTTTGGCTAGCAG (SEQ ID NO: 26)
<i>NAGS</i>	TTCGCCGTCATCGAGGTGG (SEQ ID NO: 27)	AAGAAGGCCAGGGCAAAGG (SEQ ID NO: 28)
<i>VIPR2</i>	ATCCTCCTGGGTGGGCTG (SEQ ID NO: 29)	TGGAGGTAGAGCCCCCTCC (SEQ ID NO: 30)
<i>GASI</i>	AGGAGCAGCGGCGGCG (SEQ ID NO: 31)	AAAAGAGCGGCCCAAGCAG (SEQ ID NO: 32)

It will be appreciated that many further modifications and permutations of various aspects of the described embodiments are possible. Accordingly, the described aspects are intended to embrace all such alterations, modifications, and variations that fall within the spirit and

5 scope of the appended claims.

CLAIMS

1. A method of predicting the likelihood of a liver disease in a subject, the method comprising detecting in an extracellular vesicle sample from the subject the expression of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, wherein an increase in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject is likely to have a liver disease.
2. The method of claim 1, wherein the liver disease is non-alcoholic fatty liver disease (NAFLD), non-alcoholic steatosis, or non-alcoholic steatohepatitis (NASH).
3. The method of claim 2, wherein the one or more genes is selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L* and *LIMS1*, and wherein the liver disease is non-alcoholic steatosis.
4. The method of claim 2, wherein the one or more genes is selected from the group consisting of *CTXN1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, and wherein the liver disease is non-alcoholic steatohepatitis (NASH).
5. The method of any one of claims 1 to 4, wherein the extracellular vesicle sample is an exosome sample.
6. The method of any one of claims 1 to 5, wherein the method comprises detecting a polypeptide and/or RNA product of the gene.
7. The method of claim 6, wherein the method comprises detecting an mRNA product of the gene.
8. The method of any one of claims 1 to 7, wherein the extracellular vesicle sample is isolated from a biological fluid.

9. The method of claim 8, wherein the biological fluid is blood, plasma or serum.
10. The method of claim 8 or 9, wherein the extracellular vesicle sample is isolated from at least 500 µl of the biological fluid.
11. The method of any one of claims 1 to 10, wherein the increase in gene expression is at least two-fold relative to the reference.
12. The method of any one of claims 1 to 11, further comprising isolating the extracellular vesicle sample from the subject prior to detecting gene expression.
13. A method of detecting a liver disease in a subject, the method comprising detecting in an extracellular vesicle sample from the subject the expression of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, wherein an increase in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject has a liver disease.
14. A method of treating a liver disease in a subject, the method comprising a) detecting in a sample from the subject the expression level of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, wherein an increase in the level of expression of the one or more genes, or a value derived therefrom, as compared to a reference indicates that the subject has a liver disease, and b) treating the subject who has a liver disease.
15. A method of classifying a subject with NAFLD, comprising detecting in a sample from the subject the expression level of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*,
wherein an increase in the level of expression of the one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L* and *LIMS1*, or a value derived therefrom, as compared to a reference, indicates that the subject has

non-alcoholic steatosis,

and wherein an increase in the level of expression of the one or more genes selected from the group consisting of *CTXN1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, or a value derived therefrom, as compared to a reference, indicates that the subject has non-alcoholic steatohepatitis (NASH).

16. A method of monitoring a liver disease in a subject, the method comprising detecting in a sample from the subject the expression level of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*.
17. The method of claim 16, wherein the subject is undergoing or has undergone treatment for the liver disease.
18. The method of claim 16 or 17, wherein the liver disease is non-alcoholic fatty liver disease (NAFLD).
19. The method of claim 18, wherein an increase in the level of expression of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L* and *LIMS1*, or a value derived therefrom, as compared to a reference, indicates that the subject has or is likely to have non-alcoholic steatosis.
20. The method of claim 18, wherein an increase in the level of expression of one or more genes selected from the group consisting of *CTXN1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*, or a value derived therefrom, as compared to a reference, indicates that the subject has or is likely to have non-alcoholic steatohepatitis (NASH).
21. A kit for detecting a liver disease in a subject, comprising reagents for detecting in a sample from the subject the expression of one or more genes selected from the group consisting of *CTXN1*, *CSNK1E*, *EEF1A*, *NRGN*, *CPE*, *KHDC3L*, *LIMS1*, *MADCAM1*, *CHCHD1*, *CLPS*, *HSD17B14*, *SLPI*, *CIDEB*, *NAGS*, *VIPR2* and *GAS1*.
22. The kit of claim 21, wherein the reagents comprise nucleic acids for detecting an mRNA

product of the one or more genes.

23. The kit of claim 21 or 22, further comprising devices and reagents for isolating an
exosome fraction from a biological fluid from a subject.
24. The kit of any one of claims 21 to 23, further comprising devices and reagents for
isolating mRNA from an exosome fraction.